

Osteoarthritis and Cartilage (2004) 12, 459–468

© 2004 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

doi:10.1016/j.joca.2004.02.006

Osteoarthritis and Cartilage



International
Cartilage
Repair
Society



Translation, adaptation and validation of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) for an Arab population: the Sfax modified WOMAC

Mohammad Guerhazi MD†, Serge Poiraudau MD, PhD‡,§*, Monem Yahia MD†, Monia Mezganni MD†, Jacques Fermanian MD, PhD¶, M. Habib Elleuch MD† and Michel Revel MD‡,§

† *Department of Physical and Rehabilitation Medicine, Habib Bourguiba Hospital, South University, Sfax, Tunisia*

‡ *Department of Physical and Rehabilitation Medicine, Cochin Hospital, René Descartes University, Paris, France*

§ *INSERM. Institut Fédératif de Recherche sur le Handicap (IFR 25), Paris, France*

¶ *Department of Biostatistics, Necker Hospital, René Descartes University, Paris, France*

Summary

Objective: To translate into Arabic and validate the Western Ontario and McMaster Universities (WOMAC) index.

Method: Arabic translation was obtained with use of the forward and backward translation method. Adaptations were made after a pilot study. Patients with symptomatic knee OA fulfilling the revised criteria of the American College of Rheumatology were included. Impairment outcome measures (pain as measured on a visual analog scale, the maximum distance walked, Kellgren's radiological score), Lequesne index score and Beck depression scale score were recorded. Each item was analyzed. Test–retest reliability was assessed with use of the intra-class correlation coefficient (ICC) and the Bland and Altman method. Construct validity was investigated with use of Spearman's rank correlation coefficient, and a factor analysis was performed.

Results: One hundred and three patients were included in the study. Eight questions of the WOMAC physical function subscale (PF) had insufficient psychometric properties and were excluded. Although test–retest reliability of the questionnaire was good (0.84, 0.84, and 0.92 for pain, stiffness, and modified PF subscales respectively), construct validity could not be demonstrated. Factor analysis of the modified form of the WOMAC extracted four factors, which differed from the *a priori* triple stratification. However, factor analysis of the modified PF subscales extracted two factors, which accounted for 68.4% of the total variance and could be clinically characterized (disability during activities requiring knee flexion within the first 90° and activities requiring knee flexion over more than 90°).

Conclusion: We translated and adapted the WOMAC index into Arabic to suit Tunisian people. The translated questionnaire is reliable but not valid in its original form. We propose the use of a modified version of PF subscale of the WOMAC, although the psychometric properties of this instrument must be examined in a larger population.

© 2004 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Key words: Knee, Osteoarthritis, Pain, Disability, Handicap, Translation.

Introduction

Osteoarthritis (OA) of the knee and hip are common causes of lower extremity pain and disability in the general population¹. As in western industrialized countries, in North Africa, the incidence of knee OA is more frequent than that of hip OA. An epidemiological study was recently undertaken, but results are not published yet. Pain, stiffness and physical disability are the major symptoms of OA and should be assessed when evaluating the impact of OA. Despite the high prevalence of OA², few standard methods

exist for the assessment of patients with OA either in clinical practice or in clinical trials^{3–5}.

The functional status of patients with knee OA can be assessed either by a battery of tests quantifying physical activity restrictions, such as the 6-minute walk test, the stair climbing test, and the lifting and carrying weight test⁶, or by questionnaires evaluating disability in daily living activities. This last method is relevant and appreciated for its simplicity. Moreover, it allows for assessing the patient's opinion of functional disability.

The Western Ontario and McMaster Universities (WOMAC) index is an English-language questionnaire developed and validated by Bellamy *et al.* in a comparative trial of non-steroidal anti-inflammatory drugs used for knee and hip OA⁷ and in a population with OA followed after total hip or knee arthroplasty⁸. This self-administered composite questionnaire includes five questions about pain, two about stiffness, and 17 about degree of difficulty in accomplishing daily life activities. The scores for the subscales on pain,

*Address correspondence to: Pr. Serge Poiraudau, Service de Rééducation et de Réadaptation de l'Appareil Locomoteur et des Pathologies du Rachis, 27 Rue du Fbg St Jacques, 75014 Paris, France. Tél: 01 58 41 25 49; Fax: 01 58 41 25 45; E-mail: serge.poiraudau@cch.ap-hop-paris.fr

Received 30 October 2003; revision accepted 9 February 2004.

stiffness, and disability are calculated separately. A total score, which aggregates dimensions into a global index, has been proposed but has not been validated, and the author recommends that the dimensions be kept separate and the analysis conducted on a subscale-by-subscale basis⁹. This index is progressively becoming the most widely used instrument for the assessment of OA-specific health status in Europe and North America; its metric proprieties have been extensively tested^{7,8,10}, and it is currently available in German, Swedish, Hebrew, French-Canadian, French, Italian and Spanish^{9,11–15}.

To enable comparison between assessments done in different countries, these instruments need to be not only translated but also adapted for use in different cultures^{16,17}. There is no instrument worded in the Arabic language and validated for an Arab population to measure lower limb functional disability caused by OA. Assessment is usually either by global appreciation (patient and doctor's appreciations) or use of invalidated foreign tools (often English or French).

The aim of the present study was to translate the WOMAC into Arabic and to validate the Arabic version of the index for Tunisian patients with knee OA. We expect that the translated and adapted version should suit other Arab populations because of their similar language and culture.

Materials and methods

THE SCALE

The WOMAC OA index⁹ is a self-administered composite questionnaire with a three-dimensional measure of pain, joint stiffness and degree of difficulty in accomplishing daily life activities (see annex 1 for the subscale items). The pain subscale (P subscale) includes five questions about pain for 'walking on a flat surface' (P1), 'going up or down stairs' (P2), 'at night while in bed' (P3), 'sitting or lying' (P4), and 'standing upright' (P5). The stiffness subscale (S subscale) includes two questions about stiffness 'after first wakening in the morning' (S1) and 'after sitting, lying or resting later in the day' (S2). The physical function subscale (PF subscale) includes 17 questions about the degree of difficulty when 'descending stairs' (PF1), 'ascending stairs' (PF2), 'rising from sitting' (PF3), 'standing' (PF4), 'bending to floor' (PF5), 'walking on a flat surface' (PF6), 'getting in/out of car' (PF7), 'going shopping' (PF8), 'putting on socks/stockings' (PF9), 'rising from bed' (PF10), 'taking off socks/stockings' (PF11), 'lying in bed' (PF12), 'getting in/out of bath' (PF13), 'sitting' (PF14), 'getting on/off toilet' (PF15), 'heavy domestic duties' (PF16), and 'light domestic duties' (PF17).

Each of the 24 questions is graded either on a five-point Likert scale (0 to 4) or a 100 mm visual analogue scale (VAS)⁹ ranging from 0 (no pain) to 100 (extreme pain).

The WOMAC index was translated by the forward and backward translation procedure following recommended guidelines^{16–18}. Forward translation was carried out by six professional bilingual translators (native Arab-speaking people). Translators were encouraged to strive for idiomatic rather than word-for-word translation. None of them was familiar with this type of instrument. The investigators (two physiatrists and one rheumatologist involved in OA disease) reviewed the translations to make cultural adaptations. Synthesis of the translations led to a unique version. Two other Arabic translators, one Tunisian and one Mauritanian, then carried out a backward translation of this revised version into English, to verify that the meaning

of the items was preserved. Translation difficulties, cultural diversity, conceptual equivalence and vocabulary differences were highlighted by this translation technique^{19,20}.

PRE-TESTING

For pre-testing, a sample of 13 patients with knee OA answered the translated questionnaire in order to test for misunderstanding and acceptability of the questions^{18,21–25}. This pilot study also served to determine whether a VAS or a Likert scale should be used for answers.

TESTING THE SCALE

Patients

In- and outpatients aged 40 to 80 years with symptomatic knee OA fulfilling the revised clinical and radiographic criteria of the American College of Rheumatology²⁶, for at least 6 months, were recruited from the departments of rheumatology and rehabilitation of Sfax Hospital, Tunisia. Patients were excluded if they had (a) no knee X-ray in the previous 3 months, (b) other disabling lower limb osteoarthropathy or myopathy, (c) respiratory or heart failure, (d) severe psychiatric disorders, or (e) isolated patellofemoral OA.

All patients were informed of the nature of the study and gave their consent to participate.

Methods

Demographic and clinical characteristics of patients were collected at a baseline visit.

The questionnaire was administered to 103 patients by the investigators, who could repeat the questions but were not allowed to change words. A 'never done' choice was added for questions P2, PF1, PF2, PF7, PF8, PF13, PF16, PF17 which were questions about daily activities, and instructions to patients specified that this answer was to be chosen only if the activity was not done before the appearance of symptoms of knee.

Item-by-item analysis. To check for item comprehensiveness, patients had to rate each of them as incomprehensible; doubt about understanding; comprehensible. Incomprehensible questions and questions answered 'never done' by more than 5% of patients were eliminated.

For each item a ceiling or floor effect was tested. A question was eliminated if its answer distribution was badly skewed. Test-retest reliability of each item was studied with use of the intra-class correlation coefficient (ICC)²⁷. Reliability was considered to be insufficient when the ICC was <0.65.

Psychometric properties of the scale.

Inter-rater reliability

The questionnaire was administered twice during a 24-h interval by two investigators. We chose this interval to avoid variations in clinical status and the patients remembering previous answers. Only patients who felt clinically stable were tested twice. Reliability was assessed for the P, S, and PF subscales.

Face validity

Item acceptability was studied. Item-by-item analysis was performed to detect missing responses. The time needed to complete the questionnaire was noted.

WOMAC index**استمارة وماك**

:

- | | |
|------------------------------|----------------|
| 0 : not any | 0 - لا شيء |
| 1 : a little | 1 - قليلة، |
| 2 : moderate | 2 - متوسطة، |
| 3 : important | 3 - كبيرة |
| 4 : very important - extreme | 4 - كبيرة جدا. |

P Subscale :**أ) الأوجاع:**

How much pain do you have:

ما هي شدة الأوجاع التي تحس بها ؟

- | | |
|----------------------------|--------------------------------|
| 1: walking on flat surface | 1 - عندما تمشي على أرض مستوية. |
| 2: going up or down stairs | 2 - عندما تصعد أو تنزل الدرج. |
| 3: at night while in bed | 3 - في الليل و أنت في فراشك. |
| 4: sitting or lying | 4 - عند الجلوس أو الاستلقاء. |
| 5: standing upright | 5 - عند الوقوف. |

S Subscale :

how severe is your stiffness

ب) اليبوسة أو التصلب المفصلي :

ما هي شدة اليبوسة بمفاصلك ؟

- | | |
|--|--|
| 1: After first wakening in the morning | 1 - عندما تستيقظ في الصباح. |
| 2: After sitting lying or resting later in the day | 2 - بعد الجلوس أو الاستلقاء أو الاستراحة أثناء النهار. |

PF subscale:

What degree of difficulty do you have

ج) الحركة الوظيفية :

ما هي شدة الصعوبة ؟

- | | |
|----------------------------------|--|
| 1: descending stairs | 1 - عندما تنزل الدرج. |
| 2: ascending stairs | 2 - عندما تصعد الدرج. |
| 3: rising from sitting | 3 - عند الوقوف بعد الجلوس |
| 4: standing | 4 - عند الوقوف. |
| 5: bending to floor | 5 - عندما تتحني إلى الأرض. |
| 6: walking on flat | 6 - عندما تمشي على أرض مستوية. |
| 7: getting in / out of car | 7 - عندما تركب أو تنزل السيارة. |
| 8: going shopping | 8 - عندما تذهب إلى السوق. |
| 9: putting on socks / stockings | 9 - عندما تلبس الجوارب (القليط، الكلاسط، التقاشر) |
| 10: rising from bed | 10 - عندما تقوم من الفراش. |
| 11: taking off socks / stockings | 11 - عندما تتزع الجوارب (القليط، الكلاسط، التقاشر) |
| 12: lying in bed | 12 - عندما تستلقي على الفراش. |
| 13: getting in / off bath | 13 - عندما تدخل أو تخرج من حوض الاستحمام. |
| 14: sitting | 14 - عند الجلوس على الكرسي. |
| 15: getting on / off toilet | 15 - عندما تجلس أو تقوم من المراض. |
| 16: heavy domestic duties | 16 - عندما تقوم بأعمال منزلية كبيرة. |
| 17: light domestic duties | 17 - عندما تقوم بأعمال منزلية خفيفة. |

Annex 1. English and Arab versions of WOMAC index.

Construct validity

Investigated in three ways:

- 1 Convergent validity was assessed by correlating the WOMAC subscale scores with variables that could be expected to have a converging relationship. For P

subscale, the expected correlated variables were global pain intensity in the last week assessed on a VAS (VAS P) ranging from 0 (no pain) to 100 mm (maximum pain) and the pain score of the Lequesne index (P Leq)^{4,28,29}. For S subscale, the expected correlated variable was the stiffness score of the Lequesne index

Table I
Demographic and clinical characteristics of the patients

	Mean	Standard deviation	Minimum	Maximum
Age (years)	55.9	7.67	40	78
Knee pain duration (months)	48.5	49.37	1	240
VAS pain (0 to 100 mm)	21.45	23.46	0	80
Visual analog scale score of pain (0 to 100)	49.02	25.43	0	100
Maximum distance walked (m)	1082.33	829.67	50	4000
Knee flexion (degrees)	136.02	10.78	90	140
Knee extension (degrees)	-2.04	5.66	-30	0
Lequesne score (0 to 24)	13.3	3.66	5.5	22
Kellgren's score (0 to 4)	2.74	0.77	1	4
Beck's depression score (0 to 39)	5.15	4.1	0	27

(S Leq). For PF subscale, the expected correlated variables were the functional score of the Lequesne index (F Leq) and the maximum distance walked.

- 2 Divergent validity was assessed by correlating the WOMAC subscale scores with variables known to have a weak or no relation to functional disability, stiffness and pain. These variables were the Beck depression scale score³⁰, the score of OA adapted from Kellgren's radiological score³¹ ranging from 0 (no OA) to 4 (severe OA), and the difference between the two thigh circumferences measured in centimeters at 10 cm above the patella. Moreover, for each subscale, the two other subscales were considered as divergent.
- 3 Factor analysis was performed with use of principal component analysis to extract factors. The retained factors had Eigen values >1. Independent factors were obtained with use of the varimax rotation method. Two steps were followed to study the factorial structure of the scale: The first step was to ascertain the *a priori* stratification (pain, stiffness, function) of the scale; thus, factor analysis was performed on the whole questionnaire. The second step was to assess the factorial structure of each of the subscales.

STATISTICAL ANALYSIS

Statgraphics SYSTAT 9 Delta Soft® for Windows NT/97/98 was used for all statistical analyses. Quantitative variables were described with means, standard deviations (SD), and minimum and maximum values. Qualitative variables were described with proportions and percentages.

Test-retest reliability was assessed with use of the ICC³² and the Bland and Altman method³³. These two methods give complementary information^{34,35}. The correlation between two quantitative variables was assessed with use of the non-parametric Spearman's rank correlation coefficient (r_s), because a normal distribution could not be demonstrated for all parameters studied. Spearman's rank coefficient values were interpreted as follows: excellent relationship, >0.91; good, 0.90 to 0.71; moderate, 0.70 to 0.51; fair, 0.50 to 0.31; and little or none, <0.30²⁷.

Factor analysis was performed with use of the principal analysis to extract factors. The retained factors had Eigen values >1. Independent factors were obtained with use of the varimax rotation method. The level of significance in all tests was $P < 0.05$.

Results

TRANSLATION

Forward and backward translation was carried out by eight translators and three investigators. Synthesis of the six forward translations led to a unique version. The two backward translations of this version were comparable to the original scale (annex 1).

PRE-TESTING

Patients

Thirteen patients (10 women) with a mean age of 54 ± 5.5 (minimum 46, maximum 56) of whom eight were unemployed, were given the test. Five were illiterate. The pilot study resulted in a few linguistic changes to the instrument. Terms that were incomprehensible were changed. Illiterate patients did not understand the Arabic literal translation of socks; the dialect synonyms usually used by North Africans were added to the literal ones. Some patients asked if sitting (in items P4 and PF14) meant on the floor or on a chair; for these two items we changed 'sitting' to 'sitting on a chair'. All items were retained. All patients preferred the Likert scale to the VAS. Because of the high level of illiteracy among adult Tunisian people, a self-administered questionnaire was not possible; thus, the interviewers administered the questionnaire.

TESTING THE SCALE

Demographic and clinical data

A total of 103 patients (78 women) answered the questionnaire between January and May 2002. Their mean age was 55.9 ± 7.67 years (range 40 to 78 years). Fifty-eight percent were illiterate. Table I shows the demographic and clinical characteristics of the patients.

Item analysis

Twenty-six percent of the patients had difficulty answering item PF15 and needed to know whether the question meant a Turkish toilet or a toilet with a seat. The term 'toilet' does not have the same meaning for all patients; in fact, people from the country often use a Turkish toilet, and city dwellers usually have a toilet with a seat. Therefore, item

Table II
Intra-class correlation coefficients for each question and for total score of the three subscales of the Arabic WOMAC index

Questions	P subscale	S subscale	PF subscale
1	0.80	0.87	0.62
2	0.73	0.72	0.61
3	0.72		0.73
4	0.74		0.75
5	0.72		0.70
6			0.76
7			0.86
8			0.81
9			0.72
10			0.86
11			0.69
12			0.59
13			0.81
14			0.77
15			0.77
16			0.80
17			0.64
Total score	0.84	0.84	0.92*

*Intra-class correlation coefficient value for subscale C excluding items 1, 2, 7, 12, 13, 15, 16, 17.

PF15 was eliminated. Items PF7, PF13 and PF16 were answered as 'never done' by more than 5% of the patients (6.79%, 12.61% and 18.44%, respectively) and were excluded³⁶. For each question of the scale, no ceiling or floor effect was observed.

The test-retest reliability of each question is shown in Table II. Items PF1, PF2, PF12 and PF17 had only fair repeatability, with an ICC <0.65 (0.62, 0.61, 0.59 and 0.64, respectively). These four questions were also eliminated. A modified PF subscale was obtained (annex 2).

TESTING THE MODIFIED SCALE

The test-retest reliability and construct validity were assessed after the eight items were eliminated. The score of the modified PF subscale ranged from 0 to 36.

Acceptability:

It took 5.27 ± 1.83 min (minimum 2.30 min, maximum 9 min) to complete the questionnaire. The questions were well accepted by the patients.

Inter-rater reliability

The questionnaire was administered twice to 42 patients, at an interval of 25.42 ± 11.7 h (minimum 13.5 h, maximum 49 h). The mean score at the baseline visit was 11.7 ± 3.66 (3 to 19) for P subscale, 3.69 ± 2.26 (0 to 8) for S subscale, and 14.26 ± 6.57 (2 to 27) for PF subscale. Mean scores at the second assessment were 11.95 ± 3.91 (3 to 19), 3.71 ± 1.9 (0 to 7), and 15.6 ± 6.3 (3 to 28) for P, F, and modified PF subscales, respectively. The reliability assessed by the ICC was good for P subscale (0.84), S (0.84) and modified PF subscales (0.92). Bland and

Sfax modified WOMAC

- 0 : not any
1 : a little
2 : moderate
3 : important
4 : very important - extreme

What degree of difficulty do you have

- 1: rising from sitting
2: standing
3: bending to floor
4: walking on flat
5: going shopping
6: putting on socks / stockings
7: rising from bed
8: taking off socks / stockings
9: sitting

استمارة وماك المنقحة لتقييم العجز الوظيفي بالطرف السفلي

- 0 - لا شيء
1 - قليلة،
2 - متوسطة،
3 - كبيرة
4 - كبيرة جدا.

ما هي شدة الصعوبة ؟

- 1 - عند الوقوف بعد الجلوس
2 - عند الوقوف.
3 - عندما تتحني إلى الأرض.
4 - عندما تمشي على أرض مستوية.
5 - عندما تذهب إلى السوق.
6 - عندما تلبس الجوارب (القاسيط، الكلاسط، النقاش)
7 - عندما تقوم من الفراش.
8 - عندما تنزع الجوارب (القاسيط، الكلاسط، النقاش).
9 - عند الجلوس على الكرسي.

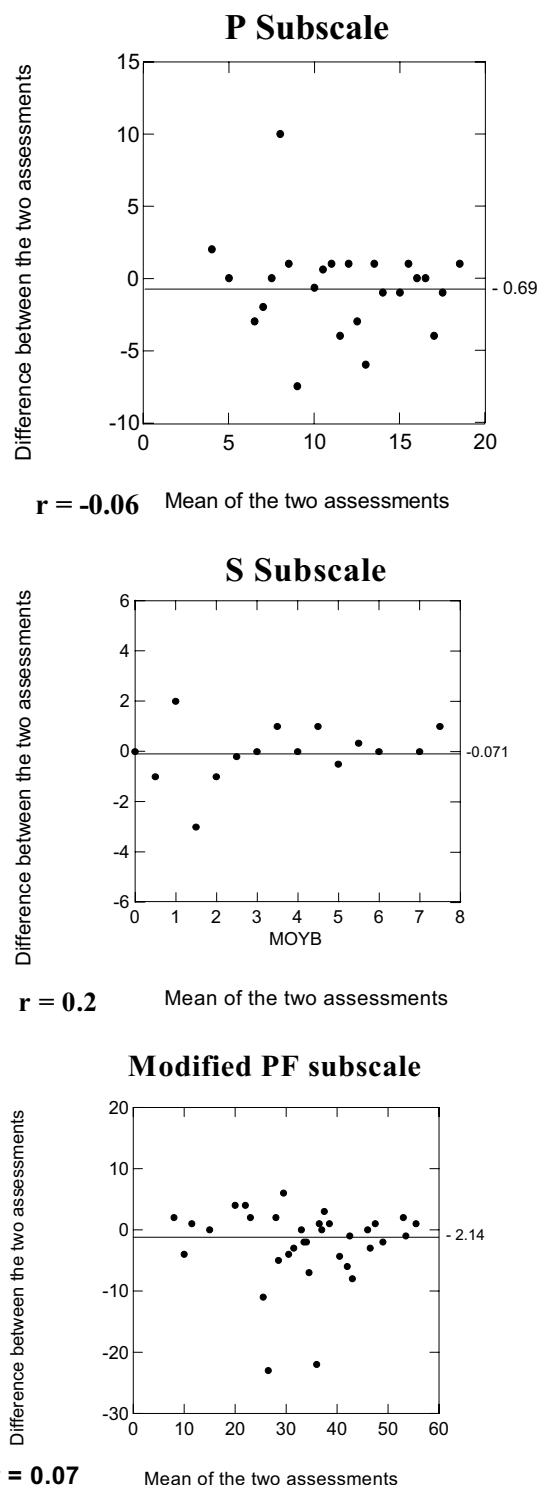


Fig. 1. Reliability of the three subscales of the Western Ontario and McMaster Universities osteoarthritis index: graphic representation according to the Bland and Altman technique; r_s : Spearman's rank correlation coefficient between differences and means of the measures.

Altman analysis revealed that the means of the differences did not differ significantly, and no systematic trend was observed (Fig. 1).

Construct validity

Table III shows the results of divergent and convergent validity for P, S, and modified PF subscales of the WOMAC index.

Convergent validity. For P subscale, expected convergent validity was observed with VAS P ($r_s=0.56$) and the pain Lequesne score (P Leq; $r_s=0.54$). For S subscale, expected convergent validity was observed with the stiffness Lequesne score (S Leq; $r_s=0.55$). For PF subscale, expected convergent validity was observed only with the functional Lequesne score (F Leq; $r_s=0.55$). The correlation with maximum distance walked was weak ($r_s=-0.32$).

Divergent validity. For each WOMAC subscale, we observed expected divergent correlations with the Beck depression scale score, Kellgren's score and the difference between the two thigh conferences. Surprisingly, correlations between subscales were fair or good. The ICC between P and S subscales was 0.57, between P and PF subscales 0.67 and between S and PF subscales 0.71.

Factor analysis

Table IV shows the results of the factor analysis of the WOMAC index and the loading of each question after varimax rotation.

Factor analysis of the whole index extracted four factors with Eigen values >1 , which differed from the *a priori* triple stratification (pain, stiffness and disability). The factors accounted for 70.5% of the total variance. Neither of the factors could be clinically characterized.

Table V shows the results of the factor analysis of P and S subscales and the modified PF subscale of the index and the loading of each question after varimax rotation.

Factor analysis of P subscale extracted one factor accounting for 50.28% of the total variance, representing knee pain, that of S subscale extracted one factor accounting for 88.26% of the total variance, representing stiffness, and that of the modified form of PF subscale extracted two factors accounting for 68.4% of the total variance. The first PF subscale factor represents disability during activities requiring knee flexion within the first 90° (items PF3, PF4, PF6, PF8, PF10, PF14) and the second, disability during activities requiring knee flexion over 90° (items PF5, PF9, PF11).

Discussion

This paper describes successive steps in translating and adapting the WOMAC index into Arabic and the psychometric properties of the translated scale for Tunisian patients with knee OA. The translated and adapted scale has good repeatability as well as each of its subscales (P, S and modified PF). However, construct validity of the translated index could not be demonstrated.

Our study showed that 58% of the patients were illiterate, which made self-administration of the questionnaire impossible. The mode of administration of a questionnaire should allow for including most of the population concerned. Like other questionnaires³⁷, our Arabic version of the WOMAC had to be administered as a hetero-questionnaire to be understood by illiterate people.

The Arabic version of the WOMAC was worded in simple and currently used literal Arabic language to allow for its

Table III
Construct validity of the three subscales of the Arabic WOMAC index (Spearman's rank correlation coefficient)

P subscale		S subscale		Modified PF subscale	
Convergent validity					
Lequesne pain score	0.54	Lequesne stiffness score	0.55	Lequesne function score	0.55
Pain on visual analog scale	0.56			Maximum distance walked	-0.32
Divergent validity					
Beck's depression scale score	0.30	Score of Beck	0.32	Score of Beck	0.30
Kellgren's radiological score	0.15	Score of Kellgren	0.24	Score of Kellgren	0.21
Circumference of the thigh	0.13	Circumference of the thigh	0.22	Circumference of the thigh	0.19
Subscale B	0.57	Subscale A	0.57	Subscale A	0.67
Subscale C	0.67	Subscale C	0.71	Subscale B	0.71

use in the largest possible Arab population. One difficulty is that literal terms are sometimes different from dialectic ones. Moreover, the dialectic synonym of a word can differ from one Arab country to another and from one region to another in the same country. When an item contains words with many dialectic synonyms, it should be eliminated, or, if it has a good clinical relevance, synonyms should be detailed in parentheses below the literal version. Thus, investigators can use the suitable version for each patient. When pre-testing the scale we found that the literal translation of 'socks' in items PF9 and PF11 was not understood by some patients. Because we considered these items clinically pertinent, we retained the two questions and provided in parentheses the dialectic synonyms usually used by North Africans.

The translation of a measurement tool can be problematic, especially if cultures differ significantly regarding socioeconomic characteristics and lifestyle³⁸. Questions considered not applicable to the target population or with an ambiguous meaning should be omitted or modified. Nevertheless, the meaning of the questionnaire must be

conserved^{37,39}. A term as simple as 'sitting' in items P4 and PF14 turned out to be ambiguous for many of our patients: difficulty and pain when sitting and when sitting on a chair can be quite different. Since the original index was validated in an occidental population, for whom sitting is understood to be on a chair rather than on the floor, and since the lifestyle of Tunisian people (and North Africans) involves both sitting on the floor and sitting on a chair we chose to add 'on a chair' in these two items.

Items PF7, PF13 and PF16 were not performed by more than 5% of patients in their daily lives. Thus, according to American Psychological Association recommendations⁴⁰, these items could not be retained. When validating the WOMAC in American, French-Canadian, German, Swedish and Hebrew populations, the authors did not point this result out, possibly because 'never done' was not offered in such validations. Faucher *et al.*^{15,41} eliminated four items from PF subscale because they were answered as 'never done' by more than 5% of patients in a French population.

Table IV
Factor analysis factors and varimax rotated factor matrix of the modified Arabic WOMAC index. The items with the highest loading are underlined

Modified WOMAC index				
Factors in factor analysis				
Factors	F1	F2	F3	F4
Eigenvalue	3.1	2.82	2.82	2.51
% variance	19.54	17.64	17.63	15.69
Cumulative %	19.54	37.18	54.81	70.5
Varimax rotated factor matrix				
Questions				
P subscale				
1	0.29	0.84	0.002	0.09
2	-0.001	0.79	0.35	0.13
3	0.67	0.14	0.15	0.04
4	0.67	0.26	0.30	0.07
5	0.31	0.17	0.77	-0.006
S subscale				
1	0.67	0.18	0.35	0.32
2	0.69	0.27	0.05	0.34
Modified PF subscale				
3	0.36	0.27	0.61	0.22
4	0.12	0.18	0.83	0.14
5	-0.04	0.04	0.45	0.67
6	0.33	0.72	0.24	0.20
8	0.35	0.63	0.21	0.26
9	0.28	0.22	0.08	0.83
10	0.44	0.19	0.51	0.36
11	0.35	0.32	0.01	0.81
14	0.55	0.15	0.44	0.36

Table V
Factor analysis factors and varimax rotated factor matrix of P and S subscales and modified PF subscale of the Arabic WOMAC. The items with the highest loading are underlined

	P sub-scale	S sub-scale	Modified PF subscale	
Factors in factor analysis				
Factors	<u>F1</u>	<u>F1</u>	<u>F1</u>	<u>F2</u>
Eigenvalue	2.63	1.59	3.29	2.64
% variance	52.63	79.81	36.56	29.37
Cumulative %	52.63	79.81	36.56	65.93
Varimax rotated factor matrix				
Questions				
P subscale				
1	<u>0.73</u>			
2	<u>0.78</u>			
3	<u>0.65</u>			
4	<u>0.71</u>			
5	<u>0.74</u>			
S subscale				
1		<u>0.89</u>		
2		<u>0.89</u>		
PF subscale				
3			<u>0.80</u>	0.21
4			<u>0.81</u>	0.40
5			<u>0.30</u>	<u>0.52</u>
6			<u>0.67</u>	0.34
8			<u>0.60</u>	0.44
9			<u>0.20</u>	<u>0.91</u>
10			<u>0.70</u>	0.37
11			<u>0.22</u>	<u>0.92</u>
14			<u>0.70</u>	<u>0.40</u>

This is the only study that offered the 'never done' choice. Items eliminated in the French study were not the same as those eliminated in our study, probably because of the differences between Tunisian and French socioeconomic conditions and lifestyles.

The repeatability of the scale was excellent. It is unlikely that these results could be due to the short interval between the test and retest. Patients had to answer the questions at the beginning of the first visit. They were then asked several other questions about their family, social and professional status, and disease effects and had to answer Beck's questionnaire. When they answered for the second time, at a 24-h interval, patients might have remembered some questions but would be unlikely to remember their previous answers. For S subscale, the ICC value observed during the validation of the original questionnaire⁷ was weak (0.61). One explanation is that patients were interviewed at an interval of one week. Nevertheless, the study by Faucher *et al.*^{15,41} resulted in the same ICC (0.61) when assessing test-retest reliability at a 3-h interval. In our study the test-retest reliability of P and S subscales and the modified PF subscale was confirmed by the Bland and Altman method³³.

Construct validity of the index could not be established. Construct validity is the main criterion of validity of a questionnaire⁴⁰. Because no gold standard currently exists to assess pain and function in knee OA⁴², we used convergent and divergent validities and factor analysis. Expected convergent and divergent validity were observed for P and S subscales but were less evident for the modified PF subscale. This subscale correlated better with P subscale (assessing pain) and S subscale (assessing stiffness) than with maximum distance walked. However, expected convergent validity was observed between the modified PF subscale and the functional Lequesne score (F Leq). Faucher *et al.*^{15,41}, using Spearman Correlation Coefficient, and Ryser *et al.*⁴³, using rash analysis, reported the same high correlation between the WOMAC subscales. Faucher *et al.*^{15,41} suggested that the high correlation between P and PF subscales was due to the pain and function items addressing the same tasks. Using the same phraseology in questions assessing pain and function for the same activity seems to disrupt patients' ability to distinguish the two concepts. This observation points out the difficulty of assessing different dimensions in the same questionnaire^{15,44}.

Factor analysis of the whole WOMAC index (P, S, and modified PF subscales) extracted four factors that differed from the *a priori* triple stratification. The concept of an algofunctional assessment of knee OA with this index was not confirmed by factor analysis in this population. Our results of construct validity contradict those of several studies assessing the WOMAC in other languages, which conclude with a valid questionnaire. Stucki *et al.*¹¹ did not achieve factor analysis and called into question their first conclusion in a second study using rash analysis⁴³. Roos *et al.*¹² and Wigler *et al.*¹³ chose questionable convergent and divergent criteria and did not perform factor analysis. Finally, only two studies used factor analysis^{15,44}. Thumboo *et al.* and Faucher *et al.* extracted factors having pain and function items addressing the same task in the same factor, as we did. One interpretation⁴⁴ is that factors could be characterized by task whether the question addresses pain or function and that the two dimensions are closely correlated. However, because the correlation between pain and function has been shown to be weak in study of several other musculoskeletal diseases such as rheumatoid

arthritis^{45,46}, low back pain⁴⁷, neck pain⁴⁸, and hand OA⁴⁹, these conclusions remain questionable.

Nevertheless, the factorial structure of each subscale seems to be acceptable. Factor analysis of the modified PF subscale extracted two factors, which accounted for 68.4% of the total variance. These two factors represent disability during activities requiring knee flexion within the first 90° and those requiring knee flexion over 90°. Thus, it seems possible to assess each subscale separately, since their psychometric properties were demonstrated.

In conclusion, we translated and adapted an Arabic version of the WOMAC index (the Sfax modified WOMAC) to suit Tunisian people. The translated questionnaire is reliable but not valid in its original form because some questions do not apply to all individuals and because of its poor construct validity. However, P, S, and modified PF subscales seem to have sufficient psychometric properties to be used separately. As P subscale was not modified, it could be used for comparisons between pain scores in patients from different countries with different cultures. This is not the case for modified PF subscale, clearly comparisons cannot be made between the original PF subscale of the WOMAC and this modified form. The impact of reducing items on the psychometric properties of the modified subscale PF must, nevertheless, be examined in a larger population. Although the scale was validated in a Tunisian population, it could suit other Arab populations, especially North Africans. Further studies are needed to confirm such a hypothesis.

References

1. Murray CJL, Lopez AD. The global burden of disease. Geneva: World Health Organisation 1997.
2. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly: the Framingham Study. *Arthritis Rheum* 1987;30:914–8.
3. Doyle DV, Dieppe PA, Scott J, Huskisson EC. An articular index for the assessment of osteoarthritis. *Ann Rheum Dis* 1981;40:75–8.
4. Lequesne M, Mery C. European guidelines for clinical trials of new antirheumatic drugs. *EULAR Bull* 1980; 9:171–5.
5. Bellamy N, Wells G, Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. *Clin Rheumatol* 1991;10:138–43.
6. Rejeski WJ, Ettinger WH Jr, Schumaker S, James P, Burns R, Elam JT. Assessing performance-related disability in patients with knee osteoarthritis. *Osteoarthritis Cart* 1995;3:157–67.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
8. Bellamy N, Buchanan W, Goldsmith C. Validation study of WOMAC. A health status instrument for measuring clinically important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Orthop Rheumatol* 1988;1:95–108.
9. Bellamy N. WOMAC Osteoarthritis User' Guide. London, Ontario, Canada: Victoria Hospital 1995.
10. Bellamy N. Outcome measurement in osteoarthritis clinical trials. *J Rheumatol* 1995;22(Suppl 43):49–51.

11. Stucki G, Sangha O, Stucki S, Michel BA, Tyndall A, Dick W, *et al.* Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne-algofunctional index in patients with knee and hip osteoarthritis. *Osteoarthritis Cart* 1998; 6:79–86.
12. Roos EM, Klassbo M, Lohmander LS. WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Western Ontario and McMaster Universities. *Scand J Rheumatol* 1999;28:210–5.
13. Wigler I, Neumann L, Yaron M. Validation study of a Hebrew version of WOMAC in patients with osteoarthritis of the knee. *Clin Rheumatol* 1999;18:402–5.
14. Choquette D, Bellamy N, Raynauld JP. A French-Canadian version of the WOMAC osteoarthritis index. *Arthritis Rheum* 1994;37:S226.
15. Faucher M, Poiraudou S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. *Osteoarthritis Cart* 2002; 10:602–10.
16. Guillemin F, Bombardier C, Beaton D. Cross cultural adaptation of health related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417–32.
17. Guillemin F. Cross cultural adaptation and validation of health status measures. *Scand J Rheumatol* 1995; 24:61–3.
18. Perneger TL, Leplège A, Guillaud H, Ecosse E, Etter JF. COOP charts in French: translation and preliminary data on instrument properties. *Qual Life Res* 1998;7:683–92.
19. Hendricson WD, Russell IJ, Prihoda TJ, Jacobson JM, Rogan A, Bishop GD, *et al.* Development and initial validation of a dual-language English-Spanish format for the arthritis impact measurement scales. *Arthritis Rheum* 1989;32:1153–9.
20. Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, *et al.* Translating Health Status Questionnaires and evaluating their quality: the IQOLA project approach. *J Clin Epidemiol* 1998; 51:913–23.
21. Perneger TH, Leplège A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. *J Clin Epidemiol* 1999;52:1037–46.
22. Wagner A, Gandek B, Aaronson NK, Acquadro C, Alonso J, Apolone G. Cross-cultural comparisons of the content of SF 36 translations across 10 countries: results from the IQOLA project. *J Clin Epidemiol* 1998;51:925–32.
23. Gonzalez V, Stewart A, Ritter P, Lorig D. Translation and validation of arthritis outcome measures into Spanish. *Arthritis Rheum* 1995;38:1429–46.
24. Leplège A, Ecosse E, Verdier A, Perneger TV. The French SF 36 Health Survey: translation, cultural Adaptation and Preliminary Psychometric Evaluation. *J Clin Epidemiol* 1998;51:1013–23.
25. Loge J, Koasa S, Hjerstad M, Kvien T. Translation and performance of the Norwegian SF 36 Health Survey in patients with rheumatoid arthritis. Data quality, scaling assumptions, reliability and construct validity. *J Clin Epidemiol* 1998;51:1129–38.
26. Altman R. Development of criteria for the classification and reporting of osteoarthritis, classification of osteoarthritis of the knee. *Arthritis Rheum* 1986;29: 1039–49.
27. Fermanian J. Mesure de l'accord entre deux juges: cas quantitatif. *Rev Epidém et Santé Publ* 1984;32: 408–13.
28. Lequesne M. The algofunctional indices for hip and knee osteoarthritis. *J Rheumatol* 1997;24:779–81.
29. Lequesne M, Mery C, Samson M, Gerard P. Indices of severity for osteoarthritis of the hip and knee. Validation. Value in comparison with other assessment tests. *Scand J Rheumatol* 1987;65:85–9.
30. Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for primary care. *Behav Res Ther* 1997;35:785–91.
31. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;16:494.
32. Shout PE, Fleiss JL. The effects of measurement errors on some multivariable procedures. *Am J Public Health* 1977;67:1188–91.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
34. Atkinson G, Nevill A. Comment on the use of concordance correlation to assess the agreement between two variables. *Biometrics* 1997;52:775–7.
35. Lin I-kuei L, Chinchilli V. Rejoinder to the letter to the editor from Atkinson and Nevill. *Biometrics* 1997; 52:777–8.
36. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press 1989.
37. Len C, Goldenberg J, Ferraz MB, Hilario MO, Oliveira LM, Sachetti S. Crosscultural reliability of the childhood Health Assessment Questionnaire. *J Rheumatol* 1994;21:2349–52.
38. Secherst L, Fay TL, Zaidi SMH. Problems of translation in cross-cultural research. *J Cross-cultural Psychol* 1972;3:41–56.
39. Ferraz MB, Oliveira RM, Araujo PMP, Atra E, Tugwell P. Crosscultural reliability of the physical ability dimension of the Health Assessment Questionnaire. *J Rheumatol* 1990;17:813–7.
40. American Psychological Association. Standards for educational and psychological testing. Washington DC: American Psychological Association 1985.
41. Faucher M, Poiraudou S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Assessment of the test-retest reliability and construct validity of the modified WOMAC index in knee osteoarthritis. *Joint Bone Spine* 2004.
42. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; 118:622–9.
43. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using rash analysis. *Arthritis Care Res* 1999;12:331–5.
44. Thumbou J, Chew LH, Soh CH. Validation of the Western Ontario and McMaster University osteoarthritis index in Asians with osteoarthritis in Singapore. *Osteoarthritis Cartilage* 2001;9:440–6.
45. Duruöz MT, Poireaudou S, Fermanian J, Menkes CJ, Amor B, Dougados M, *et al.* Development and validation of a rheumatoid hand functional disability

- scale that assess functional handicap. *J Rheumatol* 1996;23:1167–72.
46. Poiraudau S, Lefevre-Colau MM, Fermanian J, Revel M. The ability of the Cochin rheumatoid arthritis hand functional scale to detect change during the course of disease. *Arthritis Care Res* 2000;13:296–303.
47. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, *et al.* The Quebec Back Pain Disability Scale. Measurement properties. *Spine* 1995;20:341–52.
48. Wlodyka-Demaille S, Poiraudau S, Catanzariti JF, Rannou F, Fermanian J, Revel M. French translation and validation of three functional disability scales for neck pain. *Arch Phys Med Rehabil* 2002;83:376–82.
49. Poiraudau S, Chevalier X, Conrozier T, Flippo R, Lioté F, Noël E, *et al.* Reliability, validity, and sensitivity to change of the Cochin hand functional disability scale in hand osteoarthritis. *Osteoarthritis Cart* 2001;9:570–7.
-